A Statistical Diptych:
Admissible inferences - Recurrence of symmetric Markov chains

by

Morris L. Eaton[1]
University of Minnesota
Technical Report No. 542

AMS 1980 Subject Classification: Primary 62C05, Secondary 62C10, 62C15.

Key words and phrases: Improper prior distributions, formal posterior distributions, formal Bayes rules, almost admissibility, recurrence of symmetric Markov chains, prediction, predictive distributions.

[1]This work was supported in part by the National Science Foundation Grant DMS - 88-00284

Summary

Given a parametric model and an improper prior distribution, the formal posterior distribution induces decision rules in any decision problem. The results here provide conditions under which this formal Bayes method produces admissible estimators for all bounded parametric functions when the loss is quadratic. The conditions derived are shown to be equivalent to the recurrence of a natural symmetric Markov chain (on the parameter space) generated by the model and the improper prior. The results are also used to give conditions under which formal predictive distributions are admissible decision rules in certain prediction problems.

1. Introduction:

The formal Bayes method for deriving inferential procedures occupies a significant portion of both the decision theoretic and Bayesian literature. The formal Bayes representation of estimators is a standard strategy for attempting to establish admissibility - for example, see Karlin (1958), Stein (1959, 1965), Zidek (1970), Portnoy (1971), Clevenson and Zidek (1977), Berger and Srinivasan (1978) and Brown and Hwang (1982). In the Bayesian world arguments abound which attempt to justify the use of "flat", "uninformative", or "reference" prior distributions (typically improper), and implicitly, the posterior distributions these generate - see Berger (1985) for a discussion and references. Of course, any posterior distribution allows a Bayesian to solve decision problems - one just chooses actions to minimize posterior risk.

A mathematical formulation of the formal Bayes method requires some care. Given a statistical model $P(dx|\theta)$ on a sample space $\underset{\sim}{X}$ and a σ-finite improper prior distribution ν on the parameter space Θ, the marginal measure on $\underset{\sim}{X}$,

$$(1.1)\qquad M(dx) = \int P(dx|\theta)\nu(d\theta)$$

may be badly behaved (i.e., not σ-finite). However, when M is σ-finite ($\underset{\sim}{X}$ and Θ are assumed to be Polish with their Borel σ-algebras), the formal posterior distribution on Θ, $Q(d\theta|x)$, exists and satisfies

$$(1.2)\qquad P(dx|\theta)\nu(d\theta) = Q(d\theta|x)M(dx)\ -$$

the equality means that the two measures on $\underset{\sim}{X}\times\Theta$ agree. That is, $Q(\cdot|x)$ is a probability measure for each x, and for each measurable subset $B \subseteq \Theta$, $Q(B|\cdot)$ is measurable. In addition, Q is unique in the sense that if $\tilde{Q}$ also satisfies (1.2), then there is an M-null set B_o such that $x \notin B_o$ implies $Q(\cdot|x) = \tilde{Q}(\cdot|x)$. For a discussion, see Eaton (1982); an attempt to circumvent the σ-finiteness assumption on M occurs in Hartigan (1983). Throughout this paper both ν and M are assumed to be σ-finite, so a formal posterior exists.

Given an action space A and a non-negative loss function L, a formal Bayes solution to the decision problem is any function $a(x) \in A$ which for each x satisfies

$$\int L(a,\theta)\, Q(d\theta|x) \geq \int L(a(x),\theta)\, Q(d\theta|x) \tag{1.3}$$

for all $a \in A$ (we are ignoring existence and measurability issues here). For example, if $A = R^1$ and ϕ is any bounded measurable function of θ, then

$$\hat{\phi}(x) = \int \phi(\theta)\, Q(d\theta|x) \tag{1.4}$$

is a formal Bayes estimator of ϕ when the loss function is

$$L(a,\theta) = (a - \phi(\theta))^2. \tag{1.5}$$

The results in this paper focus on the general question:

(1.6) Under what conditions on the model and the improper prior ν will the formal Bayes method produce "reasonable" decision rules?

This rather vague general question is transformed into a precise mathematical problem in

(1.7) Under what conditions on the model and ν is the estimator $\hat{\phi}$ admissible (up to ν-null sets) for all bounded measurable ϕ when the loss is (1.5)?

A main result in this paper, described in Theorem 1 below, answers (1.7) in the present generality. This result and its connection with Markov chains is described in Sections 2 through 5.

Because of the generality here, Stein's notion of almost-ν-admissibility (a-ν-a) is more appropriate than admissibility (see Section 2 for the definition of a-ν-a). Much of our discussion revolves around the transition function

$$R(d\theta|\eta) = \int_{\tilde{X}} Q(d\theta|x)\, P(dx|\eta). \tag{1.8}$$

For $B \subseteq \Theta$, $R(B|\eta)$ is the average (over X) probability assigned to B by the formal posterior $Q(\cdot|x)$, when X is sampled from $P(\cdot|\eta)$. Conditions on the behavior of R supply one answer to (1.7). For a measurable subset $C \subseteq \Theta$ satisfying $0 < \nu(C) < +\infty$, consider the class of real valued functions on Θ

$$V(C) = \{h \mid \int h^2 d\nu < +\infty,\ h \geq 0,\ h(\theta) \geq 1 \text{ for } \theta \in C\}. \tag{1.9}$$

For $h \in V(C)$, set

$$\Delta(h) = \iint (h(\theta) - h(\eta))^2\, R(d\theta \mid \eta)\, \nu(d\eta). \tag{1.10}$$

Here is a key result.

<u>Theorem 1</u>: If for each C satisfying $0 < \nu(C) < +\infty$,

$$\inf_{h \in V(C)} \Delta(h) = 0, \tag{1.11}$$

then $\hat{\phi}$ is a-ν-a for each bounded measurable ϕ.

The proof of Theorem 1, given in Sections 2 and 3, uses Blyth's method (Blyth (1951)) and an application of the Cauchy-Schwarz inequality described in Appendix I. Obviously, the function $h^* \equiv 1$ yields $\Delta(h^*) = 0$, but h^* is not in $V(C)$. Thus, condition (1.11) can be interpreted as the extent to which h^* can be approximated by functions in $V(C)$. An application of Theorem 1 to random samples from one dimensional translation families which have means appears in Section 3. This application shows that (1.11) holds so the formal Bayes estimators $\hat{\phi}$ are all a-ν-a, thus providing some justification for using Lebesgue measure as an improper prior distribution in this problem.

Because the inf in (1.11) is typically not achieved by functions in $V(C)$, the successful application of Theorem 1 depends on describing "approximate" minimizers of Δ in (1.10). This leads to the introduction of a discrete time Θ-valued Markov Chain whose transition function is $R(\cdot \mid \eta)$ defined in (1.8). To see the connection, let $K \supseteq C$ satisfy $\nu(K) < +\infty$ and let

$$V(C,K) = \{h \in V(C) \mid h(\theta) = 0 \text{ for } \theta \in K^c\}. \tag{1.12}$$

In Appendix II, a minimizer of Δ over $V(C,K)$ is characterized as a certain "hitting probability" of the Markov Chain defined by R. Using this result and letting K increase to Θ yields a connection between recurrence properties of the chain and (1.11). To be precise, let $W = (\eta, W_1, W_2, \ldots)$ be the Markov Chain which starts at η and evolves according to R. Consider the stopping time

$$\sigma_C = \begin{cases} 1^{st}\ n \geq 1 \text{ with } W_n \in C \\ +\infty \text{ if no } n \text{ exists.} \end{cases}$$

Theorem 2: For each C with $0 < \nu(C) < +\infty$,

$$\inf_{h \in V(C)} \Delta(h) = \int_C [1 - \Pr\{\sigma_C < +\infty | W_o = \eta\}]\ \nu(d\eta) \tag{1.13}$$

where W_o is the initial state of the chain.

Now, if (1.11) holds, then for each C, the integral over C in (1.13) is zero. This means that for each $\eta \in C$ (except for a ν-null set), when the chain starts at η, it returns to C w.p.1. Therefore (1.11) is equivalent to a recurrence property of W (called local-ν-recurrence here).

The technical details involving the connection between Δ and the chain W are given in Appendix II. The connection established there is valid for any ν-symmetric chain (see Appendix II for the definition of ν-symmetry), and not just for chains whose transitions functions have the form (1.8). The arguments proceed from first principles and are valid for any Polish space. Fortunately, a discussion of the rather technical matter of irreducibility has been avoided (see Nummelin (1983) for such matters). For the countable state space case when the chain is irreducible, some similar-looking results appear in Griffeath and Liggett (1982) (also see Lyons (1983)).

Even though the minimizers of Δ over $V(C,K)$ can be characterized, they remain elusive. For the case $\Theta = [0,\infty)$, a heuristic method for finding "approximate" minimizers of Δ appears in Section 5. The method is successfully applied to the one dimensional Poisson case.

An alternative criterion for the evaluation of improper prior distributions was introduced in Eaton (1982). The idea there was to regard the formal posterior distribution $Q(\cdot|x)$ as a decision rule (also called an inference in Eaton (1982)), and ask for conditions under which the decision rule is a-ν-a. This approach led to the introduction of Fair Bayes Loss Functions (see Section 6 for a brief discussion of these). The admissibility of $Q(\cdot|x)$ for a variety of such loss functions is then regarded as evidence that the improper prior ν leads to sensible inferences. It was pointed out that the prediction problem could also be viewed this way. This viewpoint is developed further here.

The problem of predicting the value of some future observable random quantity on the basis of available data has received considerable attention in the statistical literature. The time series literature is replete with derivations of minimum mean squared error predictions, while the prediction of a future response, given values of covariates, is a classical problem in linear model theory which is ordinarily attacked via mean squared error considerations. No less attention is afforded the prediction problem in the Bayesian world, although the emphasis is somewhat different. Given a probabilistic model and a prior distribution, the Bayesian solution to the prediction problem is just the conditional distribution of the quantity to be predicted, given the data and the prior. This conditional distribution is called the predictive distribution and is discussed at length in the basic text by Aitchison and Dunsmore (1975). It has been argued in the literature that prediction, as opposed to say parametric estimation, is the proper activity of statisticians--partly because prediction is often the scientific question of interest, and partly because the ability of statisticians to predict can actually be checked, unlike the popular parametric estimation-confidence set activity. For an introduction to this point of view, and further references, see Geisser (1980).

The results in Section 6 give conditions under which the formal predictive distribution obtained from ν is a-ν-a. The loss functions to which the results apply have the Fair Bayes property and are described explicitly in Proposition 6.1. For these loss functions the main result in Section 6 shows that when (1.11) holds, then the formal posterior is a-ν-a. In other words, the same condition which answers (1.7) also answers the question raised in Eaton (1982). Because the Fair Bayes estimation problems are special cases of the prediction problems, (1.11) is a sufficient condition for the a-ν-a of formal posteriors on

Θ - a result of some interest to Bayesians, since one has an explicit justification for the use of some improper priors.

Section 7 contains discussion concerning open problems and the relationship of the results here with other work on admissibility. In addition, it is pointed out that certain common groups which arise in invariant statistical problems, do not support any recurrent random walks. This fact, together with the results in Section 4, strongly suggest that the routine use of invariant prior distributions for invariant problems is suspect in such cases. However, improving decision rules by modifying invariant priors remains an open problem.

2. Preliminaries

To set notation, the sample space $\underset{\sim}{X}$ and the parameter space Θ are assumed to be Polish spaces (complete separable metric spaces) equipped with the usual σ-algebras $\underset{\sim}{B}_1$ and $\underset{\sim}{B}_2$. The available data $X \in \underset{\sim}{X}$ are assumed to have a distribution belonging to the parametric family $\{P(\cdot|\theta)|\theta \in \Theta\}$. Throughout this paper, ν denotes a σ-finite improper prior distribution on Θ ($\nu(\Theta) = +\infty$). Measurable subsets $C \subseteq \Theta$ are _ν-proper_ if $0 < \nu(C) < +\infty$.

The marginal measure on $\underset{\sim}{X}$, defined by (1.1) is assumed to be σ-finite. Thus, the formal posterior $Q(\cdot|x)$ exists (such objects are sometimes called _transition functions_), and is characterized by (1.2).

We now focus attention on the estimation of bounded measurable functions $\phi(\theta)$ when the loss is (1.5). The risk function of any estimator t of $\phi(\theta)$ is

$$(2.1) \qquad R(t,\theta) = \underset{\sim}{E}_\theta(t(X) - \phi(\theta))^2.$$

Definition 2.1 (Stein (1965)). An estimator t_o is _almost-ν-admissible_ (a-ν-a) if for any estimator t_1 which satisfies

$$R(t_1,\theta) \le R(t_o,\theta) \text{ for all } \theta,$$

the set

$$\{\theta|R(t_1,\theta) < R(t_o,\theta)\}$$

has ν measure zero.

A sufficient condition for an estimator to be a-ν-a follows. It is a variation of Blyth's condition (Blyth (1951), also see Stein (1955), Zidek (1970), Brown and Hwang (1982), and Berger (1985)). For any non-negative function $g(\theta)$ which satisfies

$$0 < \int g(\theta)\ \nu(d\theta) < +\infty,$$

the marginal measure on $\underset{\sim}{X}$

$$M_g(dx) = \int P(dx|\theta)\ g(\theta)\ \nu(d\theta)$$

is a finite measure. Further, a proper posterior distribution $Q_g(\cdot|x)$ exists and satisfies

$$P(dx|\theta)\ g(\theta)\ \nu(d\theta) = Q_g(d\theta|x)\ M_g(dx). \tag{2.2}$$

It is well known that

$$\hat{\phi}_g(x) = \int \phi(\theta)\ Q_g(d\theta|x) \tag{2.3}$$

is the Bayes estimator for ϕ when the loss function is quadratic.

For a ν-proper subset $C \subseteq \Theta$, let

$$U(C) = \{g|\int g dv < +\infty,\ g \geq 0,\ g(\theta) \geq 1 \text{ for } \theta \in C\}. \tag{2.4}$$

<u>Proposition 2.1</u>: Let t_o be an estimator for $\phi(\theta)$. If for each ν-proper V,

$$\inf_{g\in U(C)} \int [R(t_o,\theta) - R(\hat{\phi}_g,\theta)]\ g(\theta)\ \nu(d\theta) = 0, \tag{2.5}$$

then t_o is a-ν-a.

<u>Proof</u>: Assume t_o is not a-ν-a, so there exists a t_1 such that $R(t_1,\theta) \leq R(t_o,\theta)$

for all θ and

$$C_1 = \{\theta | R(t_1,\theta) < R(t_o,\theta)\}$$

has positive ν-measure. Because ν is σ-finite, there is then an $\epsilon > 0$ and a ν-proper set C_2 such that

$$C_2 \subseteq \{\theta | R(t_1,\theta) \leq R(t_o,\theta) - \epsilon\}.$$

Thus, for $g \in U(C_2)$

$$\int [R(t_o,\theta) - R(\hat{\phi}_g,\theta)]\ g(\theta)\ \nu(d\theta) \geq$$

$$\int [R(t_o,\theta) - R(t_1,\theta)\ g(\theta)\ \nu(d\theta) \geq \epsilon\nu(C_2).$$

This contradiction establishes the result. □

Because ν is σ-finite, (2.5) need only be verified for a countable number of C's. Here is a precise statement.

<u>Corollary 2.1</u>: Let $\{C_n | n=1,2,\ldots\}$ be any collection of ν-proper sets with $C_n \subseteq C_{n+1}$ and $\cup C_n = \Theta$. If (2.5) holds for each C_n, then t_o is a-ν-a.

<u>Proof</u>: A minor modification of the proof of Proposition 2.1 suffices. □

Recall that the variation distance between two probability measures, α_1 and α_2, defined on the same space, is given by

(2.6) $$||\alpha_1 - \alpha_2|| = 2\sup_B |\alpha_1(B) - \alpha_2(B)|$$

where the sup ranges over the relevant σ-algebra. Further, if λ is any measure which dominates α_1 and α_2, then

(2.7) $$||\alpha_1-\alpha_2|| = \int |p_1-p_2| d\lambda$$

where $p_i = d\alpha_i/d\lambda$ is the Radon-Nikodym derivative.

Proposition 2.2: The formal Bayes estimator $\hat{\phi}$ given by (1.4) is a-ν-a for quadratic loss, if for each ν-proper set C,

$$\text{(2.8)} \qquad \inf_{g \in U(C)} \int ||Q(\cdot|x) - Q_g(\cdot|x)||^2 \; M_g(dx) = 0.$$

Proof: Take $t_o = \hat{\phi}$ in Proposition 2.1. Then, the integrated difference of the risk functions in (2.1) is given by

$$\text{(2.9)} \quad \Delta = \int[\int\{(\hat{\phi}(x)-\phi(\theta))^2 - (\hat{\phi}_g(x)-\phi(\theta))^2\} \; P(dx|\theta)] \; g(\theta) \; \nu(d\theta)$$

$$= \int\int[(\hat{\phi}(x)-\hat{\phi}_g(x))^2 + 2(\hat{\phi}(x)-\hat{\phi}_g(x))(\hat{\phi}_g(x)-\phi(\theta))] \; Q_g(d\theta|x) \; M_g(dx)$$

$$= \int(\hat{\phi}(x)-\hat{\phi}_g(x))^2 \; M_g(dx).$$

But for each x,

$$(\hat{\phi}(x)-\hat{\phi}_g(x))^2 = (\int\phi(\theta) \; Q(d\theta|x) - \int\phi(\theta) \; Q_g(d\theta|x))^2 \le$$

$$K||Q(\cdot|x) - Q_g(\cdot|x)||^2$$

where K is the bound for ϕ. Thus when (2.8) holds, (2.5) holds so $\hat{\phi}$ is a-ν-a. □

Expressions of the forms

$$\text{(2.10)} \qquad \int||Q(\cdot|x) - Q_g(\cdot|x)||^2 \; M_g(dx)$$

have appeared elsewhere in work dealing with the approximation of formal posteriors by proper posteriors (see Stein (1963)). In some related work, Stone (1965) used the expression

$$\text{(2.11)} \qquad \int||Q(\cdot|x) - Q_g(\cdot|x)||M_g(dx)$$

for densities g to measure closeness of proper to improper posteriors (see Heath and Sudderth (1989) for a relationship of this to coherence). The difference between (2.10) and (2.11) appears to be rather important. For example, it is not hard to construct cases where the inf over U(C) of (2.10) is zero, but the inf over U(C) of (2.11) is positive. The direct verification of (2.8) is difficult in most cases and it is typical to try to bound (2.10) above by more analytically tractable (in g) expressions. This we do in the next section.

Remark 2.1: In concrete problems, one usually establishes almost admissibility and then uses a separate argument to try to obtain admissibility. For example, if one can show all finite valued risk functions are continuous and if ν assigns positive measure to all non-empty open sets, then it is clear admissibility follows from almost admissibility. □

Remark 2.2: In more general decision problems, it may be possible to use the arguments above to establish a-ν-a. Blyth's method applies to any decision problem. Thus, if one can show that an integrated risk difference in a decision problem (the general analog of (2.5)) is bounded above by a constant times (2.10), then the methods developed here will be applicable to establish a-ν-a. □

3. The Condition for Almost Admissibility

A main theorem, which provides a useful condition for a-ν-a for $\hat{\phi}$, follows. Let L_2 be the set of ν-square integrable functions. For a ν-proper set C, recall that

$$V(C) = \{h \in L_2 \mid h \geq 0,\ h(\theta) \geq 1 \text{ for } \theta \in C\}. \tag{3.1}$$

The transition function $R(\cdot|\eta)$ defined in (1.8) appears here via the measure

$$T(d\theta, d\eta) = R(d\theta|\eta)\ \nu(d\eta) = \tag{3.2}$$

$$\int Q(d\theta|x)\ P(dx|\eta)\ \nu(d\eta) = \int Q(d\theta|x)\ Q(d\eta|x)\ M(dx)$$

defined on $\Theta\times\Theta$. The final expression for T shows T is symmetric and has ν as its marginals. The discussion in Appendix II implies that

$$\Delta(h) = \iint (h(\theta)-h(\eta))^2 \, T(d\theta,d\eta) \tag{3.3}$$

described in the introduction, is finite for $h\in L_2$.

Theorem 3.1: For each ν-proper set C, assume that

$$\inf_{h\in V(C)} \Delta(h) = 0. \tag{3.4}$$

Then $\hat{\phi}$ is a-ν-a for each bounded measurable ϕ when the loss is (1.5).

Proof: We verify (2.8). For each $g\in U(C)$, Corollary A.1 in Appendix I yields

$$\int ||Q(\cdot|x) - Q_g(\cdot|x)||^2 \, M_g(dx) \le 2\Delta(\sqrt{g}). \tag{3.5}$$

Setting $h = \sqrt{g}$, when (3.4) holds, then (2.8) holds. □

Corollary 3.1: Let $\{C_n|n=1,2,\ldots\}$ be any sequence of ν-proper sets satisfying $C_n \subseteq C_{n+1}$ and $\cup C_n = \Theta$. If (3.4) holds for each C_n, then the conclusion of Theorem 3.1 holds. □

Proof: Use Corollary 2.1 and repeat the proof of Theorem 3.1.

Remark 3.1: The converse of Corollary 3.1 is valid. If (3.4) holds for all C_n, then (3.4) holds for all ν-proper C. This fact is not used here.

Example 3.1: (One-Dimensional Translation)

This example concerns the additive group $R^1 = \Theta$ and a one-dimensional translation family - say $P(\cdot|\theta)$. The improper prior Lebesgue measure typically yields a formal posterior which in turn produces formal Bayes estimators

$$(3.6) \qquad \hat{\phi}(x) = \int \phi(\theta)\, Q(d\theta | x)$$

for any bounded measurable ϕ. Under very mild conditions (such as existence of a mean), the results of this example show that $\hat{\phi}$ is a-ν-a (quadratic loss) for all such ϕ. This result is not unexpected because of the admissibility of Pitman's estimators for one-dimensional translation problems [see Stein (1959)].

For this example, take $\underset{\sim}{X} = R^k$, $\Theta = R^1$

$$P(dx|\theta) = \prod_{i=1}^{k} f(x_i - \theta)\, dx_i.$$

Thus, given θ, the data are i.i.d. from the one-dimensional translation family with density f. The improper prior for this example is Lebesgue measure $\nu(d\theta) = d\theta$. For $x \in R^k$, let

$$m(x) = \int \prod_{1}^{k} f(x_i - \theta)\, d\theta.$$

Obviously, $M(dx) = m(x)dx$ is a σ-finite measure. It is well known that

$$q(\theta|x) = \begin{cases} \prod_{1}^{k} f(x_i - \theta)/m(x) & \text{if } 0 < m(x) < +\infty \\ q_o(\theta) & \text{otherwise} \end{cases}$$

(here, $q_o(\theta)$ is a fixed symmetric density of R^1) serves as a version of the conditional density of θ given x - that is,

$$(3.7) \qquad Q(d\theta|x) = q(\theta|x)\, d\theta.$$

For $v \in R^1$, define

$$t(v) = \int q(v|x) \prod_1^k f(x_i)\ dx$$

and note that

$$(3.8) \qquad t(v) = t(-v), \int t(v)\ dv = 1.$$

A routine calculation shows that the measure $T(d\theta, d\eta)$ of Theorem 3.1 is given by

$$(3.9) \qquad T(d\theta, d\eta) = t(\theta - \eta)\ d\theta d\eta.$$

Here is a sufficient condition for (3.6) to hold.

Theorem 3.2: Let $C_n = [-n,n] \subseteq R^1$ for $n=1,2,\ldots$. If

$$(3.10) \qquad \int |\theta| t(\theta)\ d\theta < +\infty,$$

then (3.4) holds for C_n and the formal Bayes estimators of bounded measurable functions obtained from the prior $d\theta$ are a-ν-a for quadratic loss.

Proof: Fix n and define $h_m, m=1,2,\ldots$ by

$$h_m(\theta) = a_m \frac{1}{1 + \frac{\theta^2}{m^2}} = a_m h\left(\frac{\theta}{m}\right)$$

where

$$a_m = 1 + \frac{n^2}{m^2}\ .$$

Since $h_m \in V(C_n)$, it suffices to show that

(3.11) $$\lim_{m\to\infty} \Delta(h_m) = 0.$$

Now, use the symmetry of t and a change of variables to obtain

$$\Delta(h_m) = 2a_m^2 \iint \frac{\left[h\left(\eta+\frac{w}{m}\right) - h(\eta)\right]}{\frac{w}{m}} h(\eta)\ wt(w)\ d\eta dw.$$

Since h has a bounded derivative and (3.10) holds, the Dominated Convergence Theorem yields

$$\lim_{m\to\infty} \Delta(h_m) = 2\iint h'(\eta)\ h(\eta)\ wt(w)\ d\eta dw.$$

which is zero. This completes the proof. □

A sufficient condition for (3.10) to hold, expressed in terms of the density f, follows.

<u>Proposition 3.3</u>: If

$$\int |v| f(v)\ dv < +\infty,$$

then (3.10) holds.

<u>Proof</u>: We only sketch the proof. First interpret m (as a function on R^{k-1}) as the marginal density of $W = (X_2-X_1, X_3-X_1, \ldots, X_k-X_1)$, where $(X_1,\ldots,X_k)$ is the sample. Let $g(x_1|w)$ denote the conditional distribution of X_1 given W (when $\theta=0$). Then, it can be shown that

$$\int |\nu| t(\nu)\ dv = \underset{\sim}{E}\underset{\sim}{E}[(|X_1-\tilde{X}_1|)|W]$$

where X_1 and $\tilde{X}_1$ are i.i.d. (given W) with $g(\cdot|w)$ as density. Thus,

$$\int |\nu| t(\nu)\ dv \le 2\underset{\sim}{E}\underset{\sim}{E}(|X_1||W) = 2\underset{\sim}{E}|X_1|. \ \square$$

The above result can be strengthened somewhat as follows. Suppose $X_1, \ldots, X_k$ are i.i.d. from f (with $\theta=0$) and let $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(k)}$ be the order statistic. For any $r \in \{1,2,\ldots,k\}$, it can be shown that

$$\int |\nu| t(\nu)\ dv \leq 2\underset{\sim}{E}|X_{(r)}|$$

so that if $X_{(r)}$ has a finite expectation, then Theorem 3.2 applies. This argument shows that Theorem 3.2 applies to Cauchy samples for $k \geq 3$, even though Proposition 3.3 does not apply. Of course, the techniques in this example also apply to dependent samples, as long as one can verify (3.10) for the appropriate t.

Naturally the above example can be used to provide conditions for a-ν-a when the model is a one-dimensional scale parameter model, and the improper prior is $d\theta/\theta$ on $(0,\infty)$. The details are omitted. This ends Example 3.1. □

We now return to the general case.

<u>Remark 3.2</u>: Consider any improper prior ν such that M is σ-finite and let $\phi \in L_2$. Because

$$\int [\int \phi(\theta)\ Q(d\theta|x)]^2\ M(dx) \leq \int\int \phi^2(\theta)\ Q(d\theta|x)\ M(dx) =$$

$$\int\int \phi^2(\theta)\ P(dx|\theta)\ \nu(d\theta) = \int \phi^2(\theta)\ \nu(d\theta) < +\infty,$$

the estimator

$$\hat{\phi}(x) = \int \phi(\theta)\ Q(d\theta|x)$$

is well defined a.e.(M). A routine calculation shows that (for quadratic loss),

$$\int R(\hat{\phi},\theta)\ \nu(d\theta) = \tfrac{1}{2}\int\int (\phi(\theta)-\phi(\eta))^2\ T(d\theta,d\eta)$$

is finite for $\phi \in L_2(\nu)$. This implies that $\hat{\phi}$ is a-ν-a for ϕ (quadratic loss), even though (3.4) may not hold. Thus, the generalized Bayes method always yields a-ν-a estimators for functions in L_2 when the loss is quadratic. Hence the real import of Theorem 3.1 is for bounded functions which are not in $L_2(\nu)$. □

Remark 3.3: Assume that (3.4) holds for a particular improper prior ν. Consider another improper prior ν_1 given by

$$\nu_1(d\theta) = \Psi(\theta)\ \nu(d\theta)$$

where Ψ is uniformly bounded away from zero and infinity - that is, there are constants c_1 and c_2 such that

$$0 < c_1 \le \Psi(\theta) \le c_2 < +\infty, \text{ all } \theta. \tag{3.12}$$

Let $\Delta(h)$ be given by (3.3) when the improper prior is ν and let $\Delta_1(h)$ be given by (3.3) when the improper prior is ν_1. It is not hard to show that

$$\Delta_1(h) \le \frac{c_2^2}{c_1^2}\,\Delta(h) \tag{3.13}$$

for $h \in V(C)$. Thus, when (3.4) holds for ν, it holds for ν_1. For example, in the translation problem of Example 3.1, we obtain a-ν-a for any prior $\Psi(\theta)d\theta$ as long as Ψ satisfies (3.12) and (3.4) holds for Lebesgue measure. □

4. The Markov chain connection:

The condition for a-ν-a given in Theorem 3.1 involves the behavior of the transition function $R(d\theta|\eta)$ defined by (3.1). This condition is

$$\begin{cases} \inf_{h \in V(C)} \Delta(h) = 0 \\ \text{for each } \nu\text{-proper set } C. \end{cases} \tag{4.1}$$

Typically, the inf in (4.1) is not achieved by a function in $L_2(\nu)$, but the inf can be approximated in the following manner. Fix a ν-proper set C and let K be ν-proper with $K \supseteq C$. With V(C,K) given by (1.12), let

$$(4.2) \qquad \delta_K = \inf_{h \in V(C,K)} \iint (h(\theta)-h(\eta))^2 \, R(d\theta|\eta) \, \nu(d\eta).$$

The main results in Appendix II provide both a formula for δ_K and a characterization of a function in V(C,K) which achieves the inf in (4.2). A statement of these results is conveniently given in the language of Markov chains.

The transition function $R(\cdot|\eta)$ defines a Markov-chain

$$W = (\eta, W_1, W_2, \ldots)$$

on the infinite product space Θ^∞ (see Neveu (1964), Chapter 5). The initial state of the chain is $W_o = \eta$ and successive states, say W_{i+1}, are generated from the probability measure $R(\cdot|W_i)$, $i=0,1,\ldots$. The probability measure of W on Θ^∞ is denoted by

$$(4.3) \qquad S(\cdot|W_o = \eta)$$

where W_o is the initial state of the chain. Observe that the chain W is <u>ν-symmetric</u> - that is, the measure

$$T(d\theta, d\eta) = R(d\theta|\eta) \, \nu(d\theta)$$

introduced in Section 3 is a symmetric measure on $\Theta \times \Theta$. This property underlies all of the results in Appendix II.

To describe a minimizer in (4.2) introduce two stopping times:

$$
(4.4)\qquad \tau = \begin{cases} 1^{st}\ n \geq 0 \text{ such that } W_n \in C \cup K^C \\ +\infty \text{ if no } n \text{ exists} \end{cases}
\qquad
\sigma = \begin{cases} 1^{st}\ n \geq 1 \text{ such that } W_n \in C \cup K^C \\ +\infty \text{ if no } n \text{ exists} \end{cases}
$$

and let $B_\tau = \{\tau < +\infty\}$, $B_\sigma = \{\sigma < +\infty\}$. Now, start the chain at $W_o = \eta$ and let $h_o(\eta)$ be the probability that the stopped chain W_τ is in C (and it stops). In symbols,

$$(4.5)\qquad h_o(\eta) = S\{(W_\tau \in C) \cap B_\tau | W_o = \eta\}.$$

Since h_o is one on C, zero on K^c and is bounded by one on $C^c \cap K$, it is in V(C,K).

<u>Theorem 4.1</u>: The function h_o in (4.5) achieves the inf in (4.2). Furthermore,

$$(4.6)\qquad \delta_K = \int_C [1 - P\{(W_\sigma \in C) \cap B_\sigma | W_o = \eta\}]\ \nu(d\eta).$$

<u>Proof</u>: See Theorem A.1 in Appendix II.

Theorem 4.1 contains important qualitative information concerning the form of functions which are "approximate" minimizers of $\Delta(h)$. However, even in the simplest examples, the explicit calculation of h_o seems hopeless, but the "rough method" described in the next section does provide some hope for finding reasonable approximations to h_o in non-trivial examples. The next result shows that the approximate minimization problem (involving K) actually converges to the minimization problem of interest when K increases to Θ.

Given a ν-proper set C, define a stopping time σ_C by

$$\sigma_C = \begin{cases} 1^{st}\ n \geq 1 \text{ such that } W_n \in C \\ +\infty \text{ if no } n \text{ exists.} \end{cases}$$

Let K_m be an increasing sequence of ν-proper sets such that $C \subseteq K_1$ and $K_m \to \Theta$.

Theorem 4.2: The following equalities hold:

$$(4.7)\quad \begin{cases} \text{(i)}\ \lim_{m\to\infty} \delta_{K_m} = \inf_{h\in V(C)} \Delta(h) \\ \text{(ii)}\ \inf_{h\in V(C)} \Delta(h) = \int_C [1 - P\{\sigma_C < +\infty | W_o = \eta\}]\ \nu(d\eta). \end{cases}$$

Proof: These are proved in Appendix II.

Now, we interpret (4.7)(ii) when the condition (4.1) for a-ν-a holds; that is, when

$$(4.8)\quad \begin{cases} \int_C [1 - P\{\sigma_C < +\infty | W_o = \eta\}]\ \nu(d\eta) = 0 \\ \text{for each } \nu\text{-proper set } C \end{cases}$$

Given the definition of local-ν-recurrence in Appendix II, we have

Theorem 4.3: The condition (4.1) for a-ν-a holds iff the symmetric chain W is locally-ν-recurrent.

Example 4.1: Take $\underset{\sim}{X} = \Theta = \{0,1,2,\ldots\}$ and let c denote counting measure on $\underset{\sim}{X}$. Consider the model with density (with respect to c)

$$f(x|\theta) = \begin{cases} p(\theta) & \text{if } x = \theta \\ 1-p(\theta) & \text{if } x = \theta + 1 \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p(\theta) < 1$ for all $\theta \in \Theta$. Let the prior distribution on Θ be $\nu(d\theta) = \pi(\theta)\ c(d\theta)$ with $\pi(\theta) > 0$.

Setting

$$m(x) = \int f(x|\theta)\ \nu(d\theta)$$

and calculating the transition function

$$R(d\theta|\eta) = r(\theta|\eta)\ \nu(d\theta),$$

we find that the transition density r is given by

$$r(0|0) = \frac{p^2(0)}{m(0)} + \frac{(1-p(0))^2}{m(1)}$$

$$r(1|0) = \frac{p(1)(1-p(0))}{m(1)}$$

and for $\eta \geq 1$,

$$r(\eta-1|\eta) = \frac{(1-p(\eta-1))p(\eta)}{m(\eta)}$$

$$r(\eta|\eta) = \frac{p^2(\eta)}{m(\eta)} + \frac{(1-p(\eta))^2}{m(\eta+1)}$$

$$r(\eta+1|\eta) = \frac{p(\eta+1)(1-p(\eta))}{m(\eta+1)} .$$

For other values of θ, $r(\theta|\eta) = 0$. Thus the chain is an irreducible random walk so recurrence and local-ν-recurrence are equivalent. Applying the well-known condition for recurrence in a random walk (see Karlin and Taylor (1975)), p.108) we find that (4.1) holds iff

(4.9) $$\sum_0^\infty \frac{1}{\pi(\theta)p(\theta)(1-p(\theta))} = +\infty.$$

In particular, if the $p(\theta)$ are uniformly bounded away from 0 and 1, (4.8) holds iff the sum of the $\pi^{-1}(\theta)$'s diverges. This supports the well-known admonition that one should not use improper priors which "put too much mass on remote portions of the parameter space." However, given <u>any</u> sequence $\pi(\theta) > 1$ with $\pi(\theta)\to\infty$ as $\theta\to\infty$, the model with $p(\theta) = \pi^{-1}(\theta)$ satisfies (4.9). Thus conditions implying (4.1) will necessarily involve both the improper prior and the model. □

The connection with Markov chains has direct implications for the use of Haar measure as an improper prior distribution when the parameter space in question is a group. For example, suppose $\underset{\sim}{X} = \Theta = R^p$ and the model is

$$P(dx|\theta) = f(x-\theta)\ dx.$$

Thus, we have one observation X from a translation family on R^p. Taking the improper prior to be the translation invariant measure on R^p, namely $\nu(d\theta)=d\theta$, a routine calculation shows that the transition function is

$$R(d\theta|\eta) = r(\theta-\eta)\ d\theta$$

where

$$r(v) = r(-v) = \int f(x-v)f(x)dx.$$

Example 3.1 shows that when $p = 1$ and

$$\int |v| r(v)\ dv < +\infty,$$

then the Markov Chain (random walk) on R^1 defined by $R(d\theta|\eta)$ is recurrent (recurrence and almost ν-recurrence are equivalent). When p=2, it is known that if

$$\int ||v||^2 r(v)\ dv < +\infty,$$

then the random walk on R^2 is recurrent. Hence the improper prior $d\theta$ on R^2 produces a-ν-a estimators. But, for $p \geq 3$, there are no non-trivial recurrent random walks on the group R^p, and thus (4.1) must fail to hold (see Guivarc'h, Keane and Roynette (1977)). Other invariant problems are discussed briefly in Section 7.

5. A Heuristic Method

The results of Section 4 characterize the minimizer of $\Delta(h)$ over the class $V(C,K)$. Typically one can calculate neither the minimizer nor the minimum explicitly. The method presented here, for the special case that $\Theta = [0,\infty)$, consists of

(5.1)
- (i) trying to bound $\Delta(h)$ by a constant times $\rho(h)$ (which is defined in (5.4))
- (ii) obtaining an explicit minimizer of $\rho(h)$ over a subclass of $V(C,K)$ for nice sets C and K
- (iii) using (ii) to derive verifiable conditions that drive $\rho(h)$ (and we hope $\Delta(h)$) to zero.

Until further notice, $\Theta = [0,\infty)$, $C = [0,a]$ and $K = [0,b]$ with $b > a > 0$. Assume h is differentiable and write

$$(h(\theta) - h(\eta))^2 = (h'(\xi))^2(\theta-\eta)^2 \tag{5.2}$$

where ξ is between θ and η. Next, replace $(h'(\xi))^2$ by what one hopes is an upper bound - namely

$$D[(h'(\theta))^2 + (h'(\eta))^2], \tag{5.3}$$

where D is a constant (not depending on b). Then, set

$$\rho(h) = \iint[(h'(\theta))^2 + (h'(\eta))^2](\theta-\eta)^2 \; R(d\theta|\eta) \; \nu(d\eta). \tag{5.4}$$

The symmetry of the measure $R(d\theta|\eta)\,\nu(d\eta)$ yields

$$\rho(h) = 2\int(h'(\eta))^2 \; \sigma(\eta) \; \nu(d\eta) \tag{5.5}$$

where

(5.6) $$\sigma(\eta) = \int(\theta-\eta)^2 \, R(d\theta|\eta).$$

Now, assume $\nu(d\eta) = \pi(\eta)\, d\eta$ and define h_b as follows:

(5.7) $$h_b(\theta) = \begin{cases} 1 \text{ if } \theta \in [0,a] \\ 0 \text{ if } \theta \in [b,\infty) \\ \dfrac{\int_\theta^b [\sigma(\eta)\pi(\eta)]^{-1} d\eta}{\int_a^b [\sigma(\eta)\pi(\eta)]^{-1} d\eta}, \ \theta \in (a,b) \ . \end{cases}$$

Of course, it is assumed that for sufficiently large a,

$$0 < \int_\theta^b [\sigma(\eta)\pi(\eta)]^{-1} d\eta < +\infty$$

for all $\theta \in (a,b)$ and all $b > a$. This choice of h_b is prompted by the fact that h_b minimizes $\rho(h)$ over those h's in V(C,K) which are a.e. differentiable and satisfy $h(a) = 1$, $h(b) = 0$. Further,

(5.8) $$\rho(h_b) = \frac{2}{\int_a^b [\sigma(\eta)\pi(\eta)]^{-1} d\eta} \ .$$

The above discussion yields

Theorem 5.1: With h_b defined by (5.7) assume that

(i) $\Delta(h_b) \le D\rho(h_b)$ for all sufficiently large b where D is a fixed constant

(ii) $\lim_{b\to\infty} \int_a^b [\sigma(\eta)\pi(\eta)]^{-1} d\eta = \infty.$

Then the condition (4.1) for a-ν-a holds.

Proof: Obvious from (5.8).

Example 5.1: Take $\underset{\sim}{X} = \{0,1,2,\ldots\}$, $\Theta = [0,\infty)$ and suppose X is Poisson with parameter θ. Consider priors of the form $\nu(d\theta) = \theta^\alpha d\theta$ where α is a parameter. In order that the marginal measure M(dx) be σ-finite it is necessary and sufficient that $\alpha \in (-1,\infty)$, which we assume. The transition function $R(d\theta|\eta)$ is

$$R(d\theta|\eta) = r(\theta|\eta)\ \nu(d\theta)$$

where

$$(5.9) \qquad r(\theta|\eta) = \exp[-\theta-\eta] \sum_{j=0}^{\infty} (\theta\eta)^j/j!\Gamma(j+\alpha+1).$$

From this, we calculate that

$$(5.10) \qquad \sigma(\eta) = \int(\theta-\eta)^2\ R(d\theta|\eta) = 2\eta + (\alpha+1)(\alpha+2).$$

Thus condition (ii) of Theorem 5.1 holds for $\alpha \in (-1,0]$, but not for $\alpha>0$. Condition (i) holds with D = 1, but is more difficult to verify. However, the argument is little more than calculus and the fact that for $\alpha \in (-1,0]$, $\theta^\alpha[\theta+(\alpha+1)(\alpha+2)]$ is increasing on $[a,\infty)$ for a large enough. The details are omitted. Thus for the Poisson, the argument shows that for $\alpha \in (-1,0]$, the improper prior $\theta^\alpha d\theta$ yields a-ν-a estimators for bounded measurable functions.

Remark 5.1: Conditions for the recurrence of Markov chains on $[0,\infty)$ were discussed in Lamperti (1960). His conditions involved

$$\mu(\eta) = \int_0^\infty(\theta-\eta)\ R(d\theta|\eta)$$

and

$$\sigma(\eta) = \int_0^\infty (\theta-\eta)^2 \; R(d\theta|\eta).$$

Lamperti showed that if

$$\mu(\eta) \le \frac{\sigma(\eta)}{2\eta} + O(\eta^{-1-\delta}) \tag{5.11}$$

for some $\delta > 0$, then the chain generated by R is recurrent. For the Poisson example, $\mu(\eta) = \alpha+1$ when the prior is $\theta^\alpha d\theta$, and $\sigma(\eta)$ is given in (5.10). Thus for $\alpha \in (-1,0)$, (5.11) holds, but not for the boundary case $\alpha = 0$. □

Conditions resembling (ii) in Theorem 5.1 have appeared elsewhere in the decision theoretic literature, typically in papers dealing with estimation of unbounded functions when the loss is quadratic (see Karlin (1958), Brown and Hwang (1982)). However, the explicit use of $\sigma(\eta)$ in this condition appears to be new. Two-sided versions of the condition when $\Theta = R^1$ also appear in some of these works.

Stein (1965) also indicated that some multidimensional problems might be amenable to arguments similar to those above. This we illustrate with a simple example when $\Theta = R^p$ (such as is the case for p-dimensional translation problems). Given a model $P(dx|\theta)$, consider a prior of the form

$$\nu(d\theta) = \xi(du) \; \pi(t) \; dt$$

where $\theta = tu$ with $t \ge 0$ and u a unit vector in R^p, so $||\theta|| = t$. Here, ξ is assumed to be a probability measure on unit vectors, so the "improper part" of the prior ν is $\pi(t)dt$ on $[0,\infty)$. (It is possible to let ξ depend on t in what follows, but we eschew that generalization). Define the new probability model

$$\tilde{P}(dx|t) = \int P(dx|tu) \; \xi(du)$$

with parameter space $[0,\infty)$. It is a routine argument to show that if (4.1) holds for the model $\tilde{P}(dx|t)$ and prior $\pi(t)dt$, then (4.1) holds for the model $P(dx|\theta)$ and prior $\nu(d\theta)$. Of course, Theorem 5.1 may apply to the $\tilde{P}$-π problem.

An alternative approach to multidimensional problems is the following obvious extension of the heuristic method described for $[0,\infty)$. Just replace (5.2) to (5.5) with the obvious multidimensional versions to obtain

$$(5.12) \qquad \tilde{\rho}(h) = 2\int ||\nabla h(\eta)||^2 \, \sigma(\eta) \, \nu(d\eta)$$

where ∇h denotes the gradient vector and

$$(5.13) \qquad \sigma(\eta) = \int ||\theta-\eta||^2 \, R(d\theta|\eta).$$

Next attempt to minimize $\tilde{\rho}(h)$ over a suitable class of h's (for nice sets C and K). Expressions similar to (5.12) have arisen elsewhere - see Brown (1971) and Srinivasan (1981) for example.

It should be mentioned that both (1.7) and the sufficient condition (1.11) are invariant under one-to-one bimeasurable transformations of Θ, as is the condition for local-ν-recurrence of the Markov chain given in Section 4. However, the heuristics proposed above are not invariant under such transformations, so it becomes relevant to ask for a "good" coordinate system in which to try the heuristics. Currently, I do not have a plausible suggestion.

6. The Prediction Problem

It is demonstrated here that the conditions for a-ν-a (4.11) imply that the posterior distribution $Q(d\theta|x)$ can be viewed directly as an admissible decision rule in certain decision problems. We focus on the prediction problem which includes the above situation as a special case. It is assumed that the reader is familiar with Eaton (1982) in what follows.

The prediction problem consists of data $X \in \underset{\sim}{X}$, a variable to be predicted $Z \in \underset{\sim}{Z}$, and an unknown parameter $\theta \in \Theta$ which indexes the probability model describing the joint distribution of X and Z. The spaces $\underset{\sim}{X}$, $\underset{\sim}{Z}$, and Θ are assumed to be Polish and the σ-algebras are those generated by the open sets. The probability model is written

$$(6.1) \qquad P(dx|z,\theta) \, S(dz|\theta)$$

where $P(\cdot|z,\theta)$ is the conditional distribution of X given z and θ, and $S(\cdot|\theta)$ is the conditional distribution of Z given θ. The marginal distribution of X given θ is then

$$(6.2) \qquad P(dx|\theta) = \int_{\underset{\sim}{Z}} P(dx|z,\theta)\ S(dz|\theta).$$

Our formulation of the prediction problem resembles that in Aitchison and Dunsmore (1975). After seeing the data X = x, one wants to specify a distribution for Z. Thus a "decision" consists of a distribution $\delta(\cdot|x)$ defined on the Borel sets of $\underset{\sim}{Z}$. In a decision theoretic framework, this means that the appropriate action space for the prediction problem is the set of all probability measures on the Borel sets of $\underset{\sim}{Z}$ - say $\underset{\sim}{M}(\underset{\sim}{Z})$. (The σ-algebra for $\underset{\sim}{M}(\underset{\sim}{Z})$ is that generated by the weak topology on $\underset{\sim}{M}(\underset{\sim}{Z})$ - see Eaton (1982) for some discussion).

A detailed argument in Eaton (1982) supports the contention that only Fair Bayes Loss Functions (see Eaton (1982) for a definition and discussion) are appropriate for the present problem. This argument hinges on the observation that if π is any proper prior distribution on Θ, the joint distribution on (X,Z,θ) given by

$$(6.3) \qquad P(dx|z,\theta)\ S(dz|\theta)\ \pi(d\theta)$$

induces a conditional distribution for Z given X = x, say $Q_\pi(dz|x)$, and this conditional distribution is the "right solution" to the prediction problem (given π). This leads to loss functions $L(a,\theta,x)$ defined on $\underset{\sim}{M}(\underset{\sim}{Z}) \times \Theta \times \underset{\sim}{X}$ which have the Fair Bayes property:

$$(6.4) \quad \begin{cases} \text{For each proper prior } \pi \text{ and for every decision rule } \delta(\cdot|x), \\ \iiint L(\delta(\cdot|x),\theta,x)P(dx|\theta,z)S(dz|\theta)\pi(d\theta) \geq \\ \iiint L(Q_\pi(\cdot|x),\theta,x)P(dx|\theta,z)S(dz|\theta)\pi(\theta). \end{cases}$$

In other words, $Q_\pi(\cdot|x)$ is a Bayes rule when the prior is π. In other contexts, loss functions (penalty functions) with this property have been discussed - see Savage (1971), Hendrickson and Buehler (1971), and Bernardo (1979) for a related notion.

Examples of loss functions which satisfy (6.4) include "quadratic forms" defined on signed measures. To be precise, let $k(\underset{\sim}{z}_1, \underset{\sim}{z}_2, x)$ be a bounded measurable function which is symmetric in $\underset{\sim}{z}_1, \underset{\sim}{z}_2$. For bounded signed measures ξ_1, ξ_2 on $\underset{\sim}{Z}$, define the bilinear function (whose dependence on x is suppressed)

$$(6.5) \qquad \langle\xi_1, \xi_2\rangle = \iint k(z_1, z_2, x)\ \xi_1(dz_1)\ \xi_2(dz_2).$$

The symmetry of k in (z_1, z_2) implies $\langle\cdot,\cdot\rangle$ is symmetric. Recall that $k(\cdot,\cdot,x)$ or $\langle\cdot,\cdot\rangle$ is <u>non-negative</u> <u>definite</u> if $\langle\xi,\xi\rangle \ge 0$ for all bounded signed measures ξ. Examples of such k's are

$$(6.6) \qquad k(z_1, z_2, x) = \sum_1^s \gamma_i(z_1, x)\ \gamma_i(z_2, x)$$

where the γ_i are bounded measurable functions.

<u>Proposition 6.1</u>: For $x \in \underset{\sim}{X}$ and $\theta \in \Theta$, let $H(\cdot|\theta, x)$ denote the conditional distribution of Z given X = x and θ. Given a bounded non-negative definite k, and hence $\langle\cdot,\cdot\rangle$, define a loss function by

$$(6.7) \qquad L(a, \theta, x) = \langle a - H(\cdot|\theta, x),\ a - H(\cdot|\theta, x)\rangle.$$

Then (6.4) holds.

<u>Proof</u>: The proof is a minor variation of arguments given in Eaton (1982). The details are omitted. □

Finally, we turn to the main problem of this section. Given an improper prior ν, assume that

$$(6.8) \qquad M(dx) = \int P(dx|\theta)\ \nu(d\theta)$$

is σ-finite. Then, as discussed in Section 2, a formal posterior distribution $\tilde{Q}(dz,d\theta|x)$ exists on $\underset{\sim}{Z} \times \Theta$ which satisfies the analogue of (1.2)

$$(6.9) \qquad P(dx|z,\theta)\ S(dz|\theta)\ \nu(d\theta) = \tilde{Q}(dz,d\theta|x)\ M(dx)$$

(obtained by replacing "θ" with "θ,z" and "$\nu(d\theta)$" with "$S(dz|\theta)\ \nu(d\theta)$"). Now, integrate out θ to obtain

$$(6.10) \qquad Q(dz|x) = \int_{\Theta} Q(dz,d\theta|x) \ -$$

the formal posterior distribution of Z given X = x, also known as the <u>formal predictive distribution</u> of Z. From our previous discussion, $Q(dz|x)$ is a decision rule for the prediction problem. The main result in this section provides sufficient conditions for the a-ν-a of the decision rule $(Q(\cdot|x)$. Indeed, it is shown that for some Fair Bayes Loss Function (those satisfying (6.4)), condition (4.1) implies a-ν-a.

<u>Remark 6.1</u>: In the special case that $\underset{\sim}{Z} = \Theta$ and $S(dz|\theta)$ is the point mass at θ, the prediction problem reduces to the "estimation" problem discussed at length in Eaton (1982) because $Q(dz|x) = Q(d\theta|x)$ is just a formal posterior. In this sense, the theory developed in the previous sections solves some of the problems posed in Eaton (1982). □

<u>Theorem 6.1</u>: Consider loss functions of the form (6.7). Assume that (4.1) holds for R based on $P(dx|\theta)$ and $\nu(d\theta)$. Then the formal predictive distribution $Q(\cdot|x)$ is a-ν-a.

<u>Proof</u>: The idea is to apply Blyth's criterion to the integrated risk difference and bound this difference above by a constant times the expression (2.10). Then results in Sections 3 and 4 apply immediately.

Because the arguments are somewhat similar to those in Eaton (1982), we only sketch the details. For a ν-proper set $C \subseteq \Theta$ and for $g \in U(C)$, (6.11) $$M_g(dx) = \int_\Theta P(dx|\theta)\ g(\theta)\ \nu(d\theta)$$

is a finite measure. The equation

(6.12) $$P(dx|z,\theta)\ S(dz|\theta)\ g(\theta)\ \nu(d\theta) = \tilde{Q}_g(dz,d\theta|x)\ M_g(dx)$$

defines the posterior distribution $\tilde{Q}_g(\cdot,\cdot|x)$ on $\underset{\sim}{Z} \times \Theta$ and the marginal posterior distribution

(6.13) $$Q_g(dz|x) = \int_\Theta \tilde{Q}_g(dz,d\theta|x).$$

Because the loss function satisfies (6.4), the integrated risk difference of concern is

(6.14) $$A(g) = \int[R(Q,\theta) - R(Q_g,\theta)]\ g(\theta)\ \nu(d\theta)$$

where, for any decision rule δ,

(6.15) $$R(\delta,\theta) = \int L(\delta(\cdot|x),\theta,x)\ P(dx|\theta).$$

Expanding (6.14) in terms of the quadratic form $<\cdot,\cdot>$ defining L, we find

(6.16) $$A(g) = \int<Q(\cdot|x) - Q_g(\cdot|x),\ Q(\cdot|x) - Q_g(\cdot|x)>M_g(dx).$$

It is now routine to bound the integrated in (6.16) above by

(6.17) $$K||\tilde{Q}(\cdot,\cdot|x) - \tilde{Q}_g(\cdot,\cdot|x)||^2$$

where K is a constant and $||\cdot||$ is variation distance. Next, use the arguments in Section 3 to obtain

$$(6.18) \qquad \int \| \tilde{Q}(\cdot,\cdot|x) - \tilde{Q}_g(\cdot,\cdot|x) \|^2 M_g(dx) \le$$

$$2\iiiint (\sqrt{g(\theta)} - \sqrt{g(\eta)})^2 \, \tilde{T}(dz_1, d\theta, dz_2, d\eta)$$

where

$$\tilde{T}(dz_1, d\theta, dz_2, d\eta) = \int \tilde{Q}(dz_1, d\theta|x) \, \tilde{Q}(dz_2, d\eta|x) \, M(dx).$$

Integrating z_1 and z_2 out in the rhs of (6.18) shows that $A(g)$ is bounded above by a constant times

$$\Delta(\sqrt{g}) = \iint (\sqrt{g(\theta)} - \sqrt{g(\eta)})^2 \, T(d\theta, d\eta)$$

where

$$T(d\theta, d\eta) = \int Q(d\theta|x) \, Q(d\eta|x) \, M(dx)$$

and $Q(d\theta|x)$ is the marginal on Θ obtained from $\tilde{Q}(dz, d\theta|x)$. Thus, when (4.1) holds, for each ν-proper set C,

$$\inf_{g \in U(C)} A(g) = 0.$$

Blyth's method then shows $Q(dz|x)$ is a-ν-a. □

7. Discussion

It is somewhat surprising that the connection between admissibility conditions and recurrence criteria is as complete as described in Section 4 - particularly given the technical issues which arose in Brown (1971) and Johnstone (1984, 1986) enroute to establishing an admissibility-recurrence connection in the normal and Poisson cases. The relationship between these two approaches is very far from clear - especially since in our approach the natural space for the Markov chain is the parameter space, while in Brown and Johnstone, the associated process is constructed on the sample space. For a discussion of related issues including an admissibility-boundary value problem tie, see Srinivasan (1981) and Johnstone (1986). Of course, the types of problems are different for at least two reasons. First, the results here give sufficient conditions for a-ν-a (quadratic loss) for the estimation of <u>all</u> bounded measurable functions, while other authors have concentrated on the estimation (quadratic loss) of a fixed "natural" parametric function (typically unbounded). Second, the sufficient conditions for admissibility in Brown (1971), Srinivasan (1981), Johnstone (1986), and others appear to be fairly close to necessary, while the necessity question is wide open here. Presumably, one natural way to phrase this necessity question is

$$(7.1) \quad \begin{cases} \text{Suppose (1.11) does not hold. Can one find a bounded parametric} \\ \text{function } \phi \text{ whose formal Bayes estimator } \hat{\phi} \text{ is not a-}\nu\text{-a for quadratic} \\ \text{loss?} \end{cases}$$

Given what is now known, it seems plausible that the earlier admissibility-recurrence connections are related to the behavior of a Markov chain on the sample space whose transition function is

$$(7.2) \qquad \tilde{R}(dx|y) = \int P(dx|\theta)\, Q(d\theta|y).$$

This chain is M-symmetric where M is the marginal measure on $\underset{\sim}{X}$ induced by ν.

The results established here do not bode well for the use of relatively invariant prior distributions when the parameter space is a non-compact Lie group - except in special circumstances. Consider a model $P(dx|\theta)$ where the parameter space Θ is a group G [for example, R^p; GTp (group of p×p lower triangular matrices with positive diagonal elements); the affine group generated by GTp and R^p], and assume the model is invariant under G (we are using the terminology and notation in Eaton (1989)). Take ν to be any relatively invariant prior distribution on G. It is fairly easy to show that the induced transition function $R(d\theta|\eta)$ on G corresponds to a random walk on G. For example, the case $G = R^p$ was discussed in Section 4. For many groups G of interest in statistics (e.g. R^p, $p \geq 3$; GTp, $p \geq 2$; the affine group generated by GTp and R^p, $p \geq 1$), the results in Guivarc'h, Keane and Roynette (1977) show that there are no non-trivial current random walks on G. Hence for these cases, (1.11) must fail and the corresponding formal posterior becomes less attractive. In invariant problems when the parameter space is a homogeneous space (rather than a group), the situation concerning random walks is less clear (see Varoupolis (1988) and Schott (1984, 1986)).

Much more work needs to be done to understand the implications of the failure of (1.11). One interesting question is the following:

$$(7.3) \quad \begin{cases} \text{Suppose (1.11) fails. Is there information in the Markov chain} \\ \text{which tells one how to modify the prior (or estimators) to produce a} \\ \text{better posterior?} \end{cases}$$

For example, if X is $N(\theta, I_p)$ with $p \geq 3$ and $\nu(d\theta) = d\theta$ on R^p, the induced transition function $R(d\theta|\eta)$ corresponds to a $N(\eta, 2I_p)$ distribution. Can one use the transience of the Markov chain to construct "improved" posterior distributions?

The criterion adopted here for the evaluation of ν is (1.7). A more stringent requirement would be to ask that the formal posterior produce a-ν-a procedures for a much wider variety of decision problems than those in (1.7). The example due to Blackwell (1951) shows some care must be taken. The following example, related to Blackwell's, shows that even in simple problems, the formal Bayes method may yield uniformly inadmissible estimators.

Example 7.1: With $\underset{\sim}{X} = \Theta = A$ = additive group of integers, suppose X given θ takes on the values θ or $\theta + 1$ each with chance 1/2. Using the flat prior on Θ, the formal posterior puts mass 1/2 at x and x - 1. It is easy to construct a bounded loss function L on $A \times \Theta$ with the following properties

(i) $L(a,a) = 0$ for all a

(ii) $L(\theta-1,\theta) < L(\theta+1,\theta)$ for all θ

(iii) $L(\theta+1,\theta) \leq L(a,\theta+1) + L(a,\theta)$ for all a, with strict inequality for $a \neq \theta+1$.

Consider the two estimators $t_o(x) = x$ and $t_1(x) = x - 1$. Using the above L, the formal Bayes method gives t_o as the unique formal Bayes estimator, but $R(t_1,\theta) < R(t_o,\theta)$ for all θ. □

Finally, it is natural to ask if the methods developed here can be adapted to give alternative proofs of standard results - for example, the exponential family results in Brown and Hwang (1982). The boundedness of the functions ϕ is used very early in the proof of Theorem 3.1. It is not clear how to adapt the material here to problems involving the estimation of unbounded functions - a standard statistical activity. In the prediction-posterior distribution problems discussed in Section 6, the boundedness of the loss function does not seem like such a bothersome assumption.

Appendix I:

A proof of inequality (3.5) follows. First, for probability measure α_1,α_2 with Radon-Nikodym derivation $p_i = d\alpha_i/d\lambda$, apply the Cauchy-Schwarz inequality to obtain a bound on variation distance:

$$\text{(A.1)} \qquad \|\alpha_1-\alpha_2\|^2 = (\int|p_1-p_2|)^2 \leq \int(\sqrt{p_1}-\sqrt{p_2})^2 \int(\sqrt{p_1} + \sqrt{p_2})^2 = 4[1 - (\int\sqrt{p_1p_2})^2].$$

An alternative bound is given in Kraft (1955).

In the notation of Sections 2 and 3, let g be a density with respect to ν so $\int g d\nu = 1$. It is easy to show M_g is absolutely continuous with respect to M. With

$$\text{(A.2)} \qquad m_g(x) = \frac{dM_g}{dM}(x) \quad ,$$

the quantity we need to bound is

(A.3) $$\delta(g) = \int ||Q(\cdot|x) - Q_g(\cdot|x)||^2\, m_g(x)\, M(dx).$$

Proposition A.1: For each density g,

(A.4) $$\delta(g) \le 2\Delta(\sqrt{g}).$$

Proof: Obviously, the set

$$A_o = \{x | 0 < m_g(x) < +\infty\}$$

satisfies $M_g(A_o^c) = 0$. The equations

$$P(dx|\theta)\, g(\theta)\, \nu(d\theta) = g(\theta)\, Q(d\theta|x)\, M(dx) = Q_g(d\theta|x)\, m_g(x)\, M(dx)$$

imply that

(A.5) $$k(x,\theta) = \begin{cases} \dfrac{g(\theta)}{m_g(x)} & \text{if } x \in A_o \\ 1 & \text{if } x \notin A_o \end{cases}$$

serves as a version of the Radon-Nikodym derivative of $Q_g(\cdot|x)$ with respect to $Q(\cdot|x)$. Now apply (A.1) with $\lambda = Q(\cdot|x)$ to get

(A.6) $$||Q(\cdot|x) - Q_g(\cdot|x)||^2 \le 4\{1 - [\int (k(x,\theta))^{1/2}\, Q(d\theta|x)]^2\}.$$

Integrating (A.6) with respect to M_g gives

(A.7) $\frac{\delta}{4} \le 1 - \int [\int \sqrt{g(\theta)}\ Q(d\theta|x)]^2\ M(dx)$

$$= 1 - \int\int (g(\theta)g(\eta))^{1/2}\ T(d\theta,d\eta)$$

$$= \frac{1}{2}\ \Delta(\sqrt{g}).$$

The last equality is a consequence of

$$\int\int g(\theta)\ T(d\theta,d\eta) = \int g(\theta)\ \nu(d\theta) = 1.$$

The proof is complete. □

Corollary A.1: For any non-negative g which satisfies

$$0 < \int g(\theta)\ \nu(d\theta) < +\infty,$$

inequality (A.4) holds.

Proof: For any $a > 0$, $\delta(ag) = a\delta(g)$ and $\Delta(\sqrt{ag}) = a\Delta(\sqrt{g})$. □

Appendix II: On Symmetric Markov Chains

In this appendix, we establish a Dirichlet principle for symmetric Markov chains which provides proofs for the assertions in Section 4. Let $(\underset{\sim}{W},\underset{\sim}{B})$ be a Polish space and let $R(\cdot|w)$ be a transition function on $\underset{\sim}{B}\times\underset{\sim}{W}$. The discrete time Markov chain on $(\underset{\sim}{W}^{\infty},\underset{\sim}{B}^{\infty})$ defined by $R(\cdot|w)$ with initial state w is denoted by $W = (w,W_1,W_2,\ldots)$. The induced probability measure for W is $\underset{\sim}{S}(\cdot|W_o=w)$ where W_o denotes the initial state of the chain.

Definition A.1: Let ν be a non-zero σ-finite measure on $(\underset{\sim}{W},\underset{\sim}{B})$. The Markov chain is ν-symmetric if the measure

(A.8) $$T(dw_1,dw_2) = R(dw_1|w_2)\ \nu(dw_2)$$

is a symmetric measure on $(\underset{\sim}{W}\times\underset{\sim}{W},\underset{\sim}{B},\underset{\sim}{B})$.

For a discussion of symmetric chains in the countable state space case, see Kelly (1979), Griffeath and Liggett (1982), and Lyons (1983). The discussion in Section 4 provides many examples of ν-symmetric Markov chains. In all that follows, W is assumed to be a ν-symmetric chain.

The following definition, a modified notion of recurrence, allows us to circumvent a discussion of irreducibility issues while relating our previous admissibility results to the recurrence of W.

<u>Definition A.2</u>: The chain W is <u>locally-ν-recurrent</u> (l-ν-r) if for each ν-proper set C, the set

$$[w \mid S(W_n \in C \text{ for some } n \geq 1 \mid W_o = w) < 1] \cap C$$

has ν-measure zero.

In words, this means that for each ν-proper C, given the chain starts in C, it returns to C w.p.1 expect for a ν-null set. Of course, when $\underset{\sim}{W}$ is countable and the chain is irreducible, l-ν-r and recurrence are equivalent.

Let $L_2(\nu)$ denote the space of ν-square integrable functions. The symmetry of T implies that

$$\iint h^2(w_1)\ T(dw_1,dw_2) = \iint h^2(w_2)\ T(dw_1,dw_2) = \int h^2(w)\ \nu(dw)$$

for all $h \in L_2(\nu)$. Hence, the Cauchy-Schwarz inequality yields

$$\left|\iint h_1(w_1)\ h_2(w_2)\ T(dw_1,dw_2)\right|^2 \leq$$
$$\int h_1^2(w)\ \nu(dw) \int h_2^2(w)\ \nu(dw)$$

for $h_1,h_2 \in L_2(\nu)$. Thus the bilinear form $\langle\cdot,\cdot\rangle$ defined by

$$\text{(A.9)}\qquad \langle h_1,h_2\rangle = \int h_1(w)\ h_2(w)\ \nu(dw) - \iint h_1(w_1)\ h_2(w_2)\ T(dw_1,dw_2)$$

is symmetric and non-negative definite for $h_1, h_2 \in L_2(\nu)$. In most of what follows, $\langle\cdot,\cdot\rangle$ is written

$$\langle h_1, h_2\rangle = (h_1, (I-R)h_2) \tag{A.10}$$

where $(\cdot,\cdot)$ is the standard bilinear form on $L_2(\nu)$ given by

$$(h_1, h_2) = \int h_1(w)\, h_2(w)\, \nu(dw),$$

I is the identity transformation, and Rh_2 is defined by

$$(Rh_2)(w) = \int h_2(w_1)\, R(dw_1|w). \tag{A.11}$$

The results in this appendix relate l-ν-r of the chain to the behavior of the form $\langle\cdot,\cdot\rangle$. To this end, let C and K_o be two ν-proper subsets of $\underset{\sim}{W}$ such that $C \subseteq K_o$. Define the stopping times τ and σ and the sets B_τ and B_σ as in (4.4). Also let

$$V(C,K_o) = \{h \in L_2(\nu) | h(w) \geq 1 \text{ for } w \in C,\ h(w) = 0 \text{ for } w \in K_o^c\}$$

and observe that

$$h_o(w) = S\{(W_\tau \in C) \cap B_\tau | W_o = w\} \tag{A.12}$$

is in $V(C,K_o)$. In fact, h_o is 1 on C and is 0 on K_o^c.

<u>Theorem A.1</u>: For a ν-symmetric chain W,

(i) $$\inf_{h\in V(C,K_o)} \langle h,h\rangle = \langle h_o, h_o\rangle,$$

and

(ii) $$\langle h_o,h_o\rangle = \int_C [1 - S\{W_\sigma \in C) \cap B_\sigma | W_o = w\}]\ \nu(dw).$$

Proof: For $h \in V(C,K_o)$, write

$$h = h_o + \phi.$$

The symmetry and non-negative definiteness of $<\cdot,\cdot>$ yields

$$<h,h> = <h_o,h_o> + 2<\phi,h_o> + <\phi,\phi>$$

$$\geq <h_o,h_o> + 2<\phi,h_o> .$$

With $Q = I\text{-}R$, (A.11) yields

$$<\phi,h_o> = \int\phi(w)(Qh_o)(w)\ \nu(dw) = \int\phi(Qh_o) =$$

$$\left(\int_C + \int_{K_o^c} + \int_{K_o\cap C^c}\right)[\phi(Qh_o).$$

The integral over K_o^c is zero because ϕ is zero on K_o^c. The integral over C is non-negative because $\phi \geq 0$ on C and $(Qh_o)(w) = 1 - (Rh_o)(w) \geq 0$ for $w \in C$. Thus

$$<\phi,h_o> \geq \int_{K_o\cap C^c} \phi(Qh_o).$$

However, a standard Markov chain argument shows that

(A.13) $\qquad (Qh_o)(w) = 0$ for $w \in K\cap C^c$

(that is, h_o is harmonic for $w \in K_o\cap A^c$). Thus (i) is established. For assertion (ii), use (A.11) and the fact that $h_o \in V(C,K_o)$ to obtain

$$<h_o,h_o> = \int h_o(Qh_o) = \int_C(Qh_o) = \int_C[1 - (Rh_o)(w)]\ \nu(dw).$$

Again a standard Markov chain argument yields

(A.14) $(Rh_o)(w) = S\{(W_\sigma \in C) \cap B_\sigma | W_o = w\}$ for $w \in C$.

This completes the proof. □

Again let C be a ν-proper set and let

$$V(C) = \{h \in L_2(\nu) | h \geq 0,\ h(w) \geq 1 \text{ for } w \in C\},$$

$$\sigma_C = \begin{cases} \text{first } n \geq 1 \text{ such that } W_n \in C \\ +\infty \text{ if no } n \text{ exists.} \end{cases}$$

<u>Theorem A.2</u>: For ν-symmetric chains,

$$\text{(A.15)} \quad \inf_{h \in V(C)} \langle h,h \rangle = \int_C [1 - P\{\sigma_C < +\infty | W_o = w\}]\ \nu(dw).$$

<u>Proof</u>: Let $\{K_m\}$ be a sequence of ν-proper sets with $C \subseteq K_1$, $K_m \subseteq K_{m+1}$ and $\underset{\sim}{W} = \bigcup_1^\infty K_m$. With $K_o = K_m$ in Theorem A.1,

$$\text{(A.16)} \quad \inf_{h \in V(C;K_m)} \langle h,h \rangle = \langle h_m,h_m \rangle = \int_C [1 - S\{W_{\sigma_m} \in C) \cap B_{\sigma_m} | W_o = w\}]\ \nu(dw)$$

where h_m and σ_m are the K_m counterparts of h_o and σ defined for K_o. Our first task is to show that

$$\text{(A.17)} \qquad \lim_{m \to \infty} \langle h_m,h_m \rangle = \int_C [1 - P\{\sigma_C < +\infty\ | W_o = w\}]\ \nu(dw)$$

To this end, let

$$E_m = \{(W_{\sigma_m} \in C)\} \cap B_{\sigma_m}$$

and let $$E = \{\sigma_C < +\infty\}.$$

Clearly $E_m \subseteq E_{m+1}$. Further, it is not hard to show that

$$E_m \to E.$$

From this and (A.16), (A.17) follows from the Dominated Convergence Theorem. Thus, the left side of (A.15) is bounded above by the right side of (A.15) because $V(C) \supseteq V(C,K_m)$ for all m.

Now, let $h \in V(C)$ and set

$$u_m = hI_{K_m} \in V(C,K_m).$$

Applying the Dominated Convergence Theorem, Theorem A.1 and (A.17) in that order yields

$$\langle h,h\rangle = \lim_{m\to\infty} \langle u_m,u_m\rangle \geq \lim_{m\to\infty} \langle h_m,h_m\rangle = \int_C [1 - P\{\sigma_C < +\infty | W_o = w\}]\ \nu(dw).$$

Thus (A.15) holds. □

Theorem A.3: The chain W is 1-ν-r iff for each ν-proper set C,

$$\inf_{h\in V(C)} \langle h,h\rangle = 0.$$

Proof: This is immediate from Theorem A.2. □

Acknowledgements: Thanks to my colleagues and friends at Minnesota, Brad Johnson, David Lane, Ashok Maitra and Bill Sudderth for comments and suggestions regarding various aspects of this work. In 1988, lectures on this material were given in Holland where conversations with Kobus Oosterhof, Frank den Hollander and Michael Keane were most helpful. Special thanks to Iain Johnstone for providing the essence of the proof of my original conjecture on the admissibility - Markov chain connection.

References

Aitchison, J. and Dunsmore, I.R. (1975). Statistical Prediction Analysis, Cambridge University Press, New York.

Berger, J. (1985). Statistical Decision Theory and Bayesian Analysis, Second Edition. Springer-Verlag, New York.

Berger, J. and Srinivasan, C. (1978). Generalized Bayes estimators in multivariate problems. Ann. Statist. 6, 783-801.

Bernardo, J. (1979). Expected information as expected utility. Ann. Statist. 7, 686-690.

Blackwell, D. (1951). On the translation parameter problem for discrete variables. Ann. Math. Statist. 18, 105-110.

Blyth, C.R. (1951). On minimax statistical procedures and their admissibility. Ann. Math. Statist. 22, 22-24.

Brown, L.D. (1971). Admissible estimators, recurrent diffusions, and insolvable boundary value problems. Ann. Math. Statist. 42, 855-904.

Clevenson, M. and Zidek, J. (1977). Simultaneous estimation of the mean of independent Poisson laws. J. Amer. Statist. Assoc. 70, 698-705.

Eaton, M.L. (1982). A method for evaluating improper prior distributions. Decision Theory and Related Topics III, Vol. 1, ed. by Berger and Gupta, 329-352.

Eaton, M.L. (1989). Group Invariance Applications in Statistics. Regional Conference Series in Probability and Statistics, Vol. 1. Published by Institute of Mathematical Statistics and American Statistical Association.

Geisser, S. (1980). A predictivistic primer. In Bayesian Analysis in Econometrics and Statistics, ed. A. Zellner, North Holland, Amsterdam, 363-381.

Griffeath, D. and Liggett, T. (1982). Critical phenomena for Spitzer's reversible nearest particle systems. Ann. Probab. 10, 881-895.

Guivarc'h, Yves, Kean, Michael and Roynette, Bernard (1977). Marches aléatoires sur les Groupes de Lie. Lecture Notes in Math. Vol. 624, Springer-Verlag, New York.

Hartigan, J. (1983). Bayes Theory, Springer-Verlag, New York.

Heath, D. and Sudderth, W. (1989). Coherent inference from improper priors and from finitely additive priors. Ann. Statist. 17, 907-919.

Hendrickson, A.D. and Buehler, R.J. (1971). Proper scores for probability forecasters. Ann. Math. Statist. 42, 1916-1921.

Johnstone, I. (1984). Admissibility, difference equations and recurrence in estimating a Poisson mean. Ann. Statist. 12, 1173-1198.

Johnstone, I. (1986). Admissible estimation, Dirichlet principles, and recurrence of birth-death chains on Z_t^p. Probab. Th. Related Fields 71, 231-269.

Karlin, S. (1958). Admissibility for estimation with quadratic loss. Ann. Math. Statist. 29, 411-415.

Kelly, F.P. (1979). Reversibility and Stochastic Networks, Wiley, New York.

Kraft, C. (1955). Some conditions for consistency and uniform consistency of statistical procedures. Univ. California Publications in Statist. Vol. 2, 125-142.

Lamperti, J. (1940). Criteria for the recurrence or transience of stochastic processes, I. J. Math. Anal. and Appl. 1, 314-330.

Lyons, T. (1983). A simple criterion for transience of a reversible Markov Chain. Ann. Probab. 11, 393-402.

Neveu, J. (1965). Mathematical Foundations of the Calculus of Probability, Holden Day, San Francisco.

Nummelin, E. (1984). General Irreducible Markov chains and Non-Negative Operators, Cambridge University Press, New York.

Portnoy, S. (1971). Formal Bayes estimation with applications to a random effects model. Ann. Math. Statist. 42, 1379-1402.

Savage, L.J. (1971). The elicitation of personal probabilities and expectations. J. Amer. Statist. Assoc. 66, 783-801.

Schott, R. (1984). Random walks on homogeneous spaces. In Probability Measures on Groups VII, ed. H. Heyer, 564-575. Lecture Notes in Mathematics, #1064. Springer-Verlag, New York.

Schott, R. (1986). Recurrent random walks on homogeneous spaces. In Probability Measures on Groups VIII, ed. H. Heyer, 146-152. Lecture Notes in Mathematics #1210. Springer-Verlag, New York.

Srinivasan, C. (1981). Admissible generalized Bayes estimators and exterior boundary value problems. Sankhya (Ser. A). 43, 1-25.

Stein, C. (1955). A necessary and sufficient condition for admissibility. Ann. Math. Statist. 76, 518-522.

Stein, C. (1959). The admissibility of Pitman's estimator of a single location parameter. Ann. Math. Statist. 30, 970-979.

Stein, C. (1965). Approximation of improper prior measures by prior probability measures. In Bernoulli-Bayes-Laplace Festschrift, 217-240. Springer-Verlag, New York.

Stein, C. (1965). Lectures notes on decision theory. Unpublished.

Stone, M. (1965). Right Haar measure for convergence in probability to quasi-posterior distributions. Ann. Math. Statist. 36, 440-453.

Varopoulos, N. Th. (1988) Green's function and harmonic functions on manifolds. In Geometry of Random Motion, ed. R. Durrett and M. Pinsky. Contemporary Mathematics, Vol. 73. American Mathematical Society, Providence, R.I.

Zidek, J.V. (1970). Sufficient conditions for admissibility under squared errors loss of formal Bayes estimators. Ann. Math. Statist. 41, 446-456.